In: Mathematical Biology Research Trends
Editor: Lachlan B. Wilson, pp. 1-19
ISBN 978-1-60456-141-8

Chapter 8

SYSTEMS BIOLOGY OF THE TRYPTOPHAN OPERON

M. Santillán[1,2]*; ***Eduardo S. Zeron***[3†] ***and M.C. Mackey***[2,4‡]
[1] Unidad Monterrey,
Centro de Investigación y Estudios Avanzados del IPN,
Av. Cerro de las Mitras No. 2565, Col. Obispado,
Monterrey, NL 64060, MÉXICO,
[2]Centre for Nonlinear Dynamics in Physiology and Medicine,
McGill University, 3655 Promenade Sir William Osler,
Montreal, QC H3G 1Y6, CANADA
[3]Departamento de Matemáticas, CINVESTAV-IPN,
Apartado Postal 14-740, México D.F. 07000, MÉXICO
[4]Departments of Physiology, Physics, and Mathematics,
McGill University, Montreal, QC, CANADA

Abstract

*E-mail address: moises.santillan@mac.com, URL: `http://homepage.mac.com/moises.santillan/`.
†E-mail address: eszeron@math.cinvestav.mx
‡E-mail address: michael.mackey@mcgill.ca

1. Introduction

The knowledge of the complete genome of a given species is just a small piece of the information thought to be useful in the understanding of one of the most complicated and important puzzles in science: "How does a biological system work?". To fully understand the behavior of an organism, an organ, or even a single cell, we need to understand the underlying gene regulatory dynamics. Given the complexity of even a single cell, answering this questions is impossible at the moment and will remain so for the foreseeable future. However, by analyzing the simplest genetic regulatory systems we may be able to develop the mathematical techniques and procedures required to tackle ever more complex genetic networks in the future.

In this paper, we review our efforts of the past few years to understand, via mathematical modeling, the dynamic behavior of one of the most studied gene regulatory networks in bacteria: the tryptophan operon. As we shall see, it is possible to obtain valuable information with relatively simple models, despite all of the assumptions underlying them.

2. A Brief Historical Review

During the Enlightenment, in the latter part of the 18th and early part of the 19th Centuries, scientific disciplines started to be hierarchically classified. This classification into the so called *exact sciences* (physics and mathematics) and those of the *life sciences* (biology and medicine) has led to the notion that these two broad divisions have evolved following distinct, separate, and sometimes even contradictory or conflicting pathways. The truth, however, is far different. Since the origins of modern science, there have been people making important contributions to both the exact and life sciences as well as the very interface between them. William Harvey discovered blood circulation with the aid of a mathematical model. Electrodynamics started with the work of Galvani and Volta (both physicians) on animal electricity. Later, Helmholtz (also a physician, but better known for his contributions to physics) invented the myograph and the ophthalmoscope, recorded for the first time the velocity of a nervous impulse, discovered the first law of thermodynamics (based on metabolic considerations), and helped to settle the foundations of all modern theories of resonance with his studies on auditory physiology.

During the 20th century, electrophysiology (the science that studies the interactions between biological tissues and electromagnetic fields) advanced enormously. Archibald V. Hill, Bernard Katz, Max Planck, Walter Nernst, Kenneth S. Cole, Alan L. Hodgkin, Andrew, F. Huxley, Erwin Neher, and Haldan K. Hartline, among others, made important contributions to its progress. Some remarkable events in the history of electrophysiology were: the explanation for the origin of the action potential, elucidated by Hodgkin and Huxley with the aid of highly sophisticated mathematical models; the Huxley cross-bridge model for muscle contraction; and the invention of the patch clamp technique by Neher.

Charles Darwin published his theory of evolution through natural selection in 1859. From the beginning, it was clear that this theory lacked proper statistical foundations, and this was its main weakness. Indeed, an apparent contradiction between Darwinism and Mendel's laws of inheritance arose immediately after the Mendelian laws were rediscovered in 1900. This gap was closed through the work of many mathematicians who, between

1860 and 1940, developed the necessary statistical tools to fuse genetics and Darwinism. The result is what we know today as population genetics or neo-Darwinism. Some of the most important contributors to this success were: Fleeming Jenkin, Francis Galton, Karl Pearson, Raphael Weldon, Godfrey H. Hardy, Ronald A. Fisher, Sewall Wright, and Theodosius Dobzhansky. Interestingly, Godfrey H. Hardy and Ronald A. Fisher are well known in the mathematical community for their contributions to real analysis and number theory [Hardy], and probability and statistics [Fisher].

Molecular biology consolidated between 1940 and 1960, and until 1970 two different schools were recognized: the structural and the informatics schools. The physicists W. H. Bragg and W. L. Bragg (father and son) founded the structural school in Cambridge. They invented X-ray crystallography in 1912, and the structural analysis of biological molecules soon started in their lab. Some of the best known structuralists were W. T. Astbury , John D. Bernal, Max Perutz, and John C. Kendrew (all of them from Cambridge), as well as Linus Pauling from Caltech. The structuralists were convinced that no new physical laws were required to explain vital phenomena. They endeavored to explain the function of biomolecules (and so of tissues and organs) from their inner structure. The secondary proteinic structure known as the alpha helix, and the structures of hemoglobin and myoglobin are some of the most important discoveries from this school.

Inspired on the uncertainty principle of quantum mechanics, Niels Bohr proposed that new principles from physics may be necessary to understand life. With this assertion, Bohr founded the informatics school of molecular biology. Max Delbrück and Erwin Schrödinger (both physicists) were two of the most important spokesmen for this school. When the Nazis took power in Germany, Delbrück moved to the USA were he started a very successful collaboration with Salvador Luria on bacteriophage research. This collaboration greatly advanced our knowledge of the molecular basis of genetics. On the other hand, Schrödinger had to move to Dublin after the Nazis invaded Austria, and there he published a little book entitled *What is life?*, which was tremendously influential on the development of molecular biology.

Some of those recruited by Schrödinger's book were James Watson and Francis Crick (who later discovered the structure of DNA in 1953), Maurice Wilkins (who provided essential physical data to Watson and Crick), Seymour Benzer (who sequenced the first gene), and François Jacob (who discovered mRNA and, together with Jacques Monod, the regulatory mechanisms of the *lac* operon.

During the second half of the 20th Century, biomathematics, also known as mathematical biology, developed as a branch of applied mathematics. Biomathematics is an active field of research and interest in it is accelerating as is the number of individuals working on it. It is essentially dedicated to mathematical modeling biological phenomena. Biomathematicians have made important contributions to ecology (through population dynamics), epidemiology, pattern formation (through the study of reaction diffusion equations), molecular biology, integrative physiology, and medicine. Some of the most best known biomathematicians are B. van der Pol, A. J. Lotka, V. Volterra, A. Turing, J. M. Smith, A. T. Winfree, etc. Readers interested in learning more about the common history of biology, mathematics, and physics are recommended to read References [1, 2, 3, 4, 5, 6].

3. Systems Biology

A new designation for an area of interdisciplinary research in biology, currently termed *systems biology*, emerged a few year ago; it continues with the long tradition described in the previous section, and especially with the long tradition of an integrative approach in physiology. The closest ancestors to what is called systems biology are systems theory and cybernetics. Since systems biology inherits part of the philosophy and the goals of both, it is interesting to briefly review these latter two.

Systems theory is an interdisciplinary field which studies relationships between systems as a whole. It was founded in the 1950s and focuses on organization and interdependence of relationships. Systems dynamics is a central part of systems theory; it provides methods for understanding the dynamic behavior of complex systems. Such methods rely on the recognition that the structure of any system—the multi-circular, interlocking, sometimes time-delayed relationships among its components—is often just as important in determining its behavior as the individual components themselves. Indeed, in many cases, it is impossible to explain the behavior of the whole system in terms of the behavior of its separated parts only. Examples are chaos theory and social dynamics.

Cybernetics is the study of communication and control, typically involving regulatory feedback, in living organisms, in machines, and in combinations of the two. It is an earlier but still-used generic term for many of the subject areas that are subsumed under the headings of adaptive systems, artificial intelligence, complex systems, complexity theory, control theory, decision support systems, dynamical systems, information theory, learning organizations, mathematical systems theory, operations research, simulation, and systems engineering.

Contemporary cybernetics began in the 1940s as an interdisciplinary study connecting the fields of control theory, electrical network theory, logic modeling, neuroscience, and human physiology. The emphasis of cybernetics is on the functional relations that hold between the different parts of a system; rather than the parts themselves. These relations include the transfer of information, and circular relations (feedback) that result in emergent phenomena such as self-organization. The name cybernetics was coined by its intellectual father, Norbert Wiener, to denote the study of "teleological mechanisms" (i.e. machines with corrective feedback) and was popularized through his book *Cybernetics, or Control and Communication in the Animal and Machine* (1948). Wiener was, incidently, one of the most influential and original mathematicians of the first half of the 20th century.

Systems biology is an academic field that seeks to integrate different levels of information, and so to understand how biological systems work. By studying the relationships and interactions between various parts of a biological system (e.g., gene and protein networks involved in cell signalling, metabolic pathways, organelles, cells, physiological systems, organisms, etc.) it is hoped that eventually a comprehensive model of the whole system can be developed. As the intellectual grandchild of what was originally called human physiology it endeavors to expand physiology to include biochemistry as well as molecular biology.

In contrast to much of molecular biology, systems biology does not seek to break down a system into all of its parts and to study each part of the process at a time with the hope of being able to reassemble all the parts into a whole again. Systems biology begins with the study of genes and proteins in an organism using high-throughput techniques to quantify

changes in the genome and proteome in response to a given perturbation. These techniques include microarrays to measure the changes in mRNAs and mass spectrometry, which is used to identify proteins, detect protein modifications, and quantify protein levels. However, systems biology is much more since it balances these molecular details against whole system performance and behavior. Using this integrated knowledge, the system biologist can formulate hypotheses that explain a system's behavior. Importantly, these hypotheses can be used to mathematically model the system. Models are then used to predict how different changes in the environment affect the system itself, and so they can be iteratively tested for their validity.

Recent analysis has revealed that cell signals do not necessarily propagate linearly. Instead, cellular signalling networks can be used to regulate multiple functions in a context dependent fashion. Because of the magnitude and complexity of the interactions inside the cell, it is often impossible to understand intuitively the *systems behavior* of these networks. Rather, it has become necessary to develop mathematical models and analyze the behavior of these models, both to develop a systems-level understanding and to obtain experimentally testable predictions.

New approaches to these problems are constantly being developed by quantitative scientists, such as computational biologists, statisticians, mathematicians, computer scientists, engineers, and physicists, to improve our ability to create, refine, and retest the models until the predicted behavior accurately reflects the seen phenotype.

The reader interested in reading more about the definition and philosophy of systems biology may find References [7, 8, 4, 9, 5, 10, 11, 12] appealing.

4. The Central Dogma of Molecular Biology

The central dogma of molecular biology deals with the information flow between DNA, RNA, and proteins. The standard information-flow pathway can be summarized in a very short and oversimplified manner as follows: DNA $\rightarrow$ RNA $\rightarrow$ proteins. Proteins in turn facilitate the previous two steps, as well as the replication of DNA. This whole is therefore broken down into three steps: transcription, translation, and replication. Nevertheless, information can flow backwards in some steps; see Figure 1. Below, two of the processes accounted for by the central dogma (transcription and translation) are briefly reviewed.

Transcription

Transcription is the process through which a DNA amino acid sequence is copied by an enzyme known as RNA polymerase to produce a complementary RNA. In other words, it is the transfer of genetic information from DNA into RNA. In the case of protein-encoding DNA, transcription is the beginning of the process that ultimately leads to the translation of the genetic code (via the mRNA intermediate) into a functional peptide or protein.

In prokaryotic cells, like bacteria, transcription initiation takes place through the following steps:

- RNA polymerase (RNAP) recognizes and specifically binds to a DNA segment known as the promoter. At this stage, the DNA is double-stranded and (*closed*). This RNAP/wound-DNA structure is referred to as the closed complex.

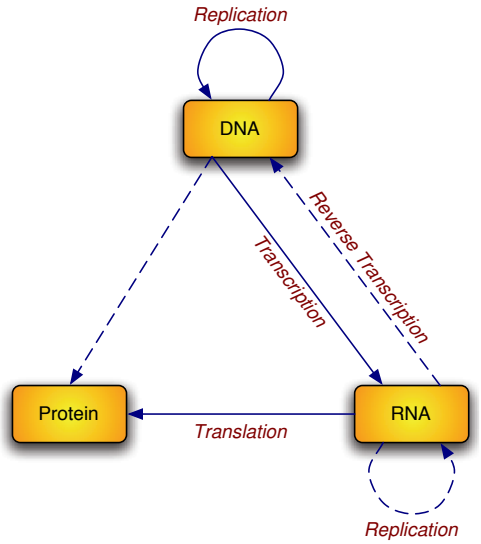

Figure 1. The 1970 version of the Central Dogma. The arrows represent the flow of information. Solid arrows represent *probable* information flow, while dotted arrows represent *possible* information flow. Note that information flow from proteins to RNA or DNA is regarded as impossible

- The DNA is unwound and becomes single-stranded (*open*) in the vicinity of the initiation site. This RNAP/unwound-DNA structure is called the open complex.

- The RNA polymerase transcribes DNA into RNA.

Promoters can differ in *strength*; that is, how actively they promote transcription of their adjacent DNA sequence. Promoter strength is in many (but not all) cases, a matter of how tightly RNA polymerase and its associated accessory proteins bind to their respective DNA sequences. The more similar the sequences are to a consensus sequence, the stronger the binding is.

Translation

In prokaryotic cells, a nascent messenger RNA (mRNA) molecule is bound by a ribosome, where it is translated. The mRNA is read by the ribosome as triplet nucleotide sequences (codons). Complexes of initiation factors and elongation factors bring aminoacylated transfer RNAs (tRNAs) into the ribosome-mRNA complex, matching the codon in the mRNA to the anti-codon in the transfer RNA (tRNA), thereby adding the correct amino acid in the sequence encoding the gene. As the amino acids are linked into the growing peptide chain, they begin folding into the correct conformation. This folding continues until the nascent polypeptide chains are released from the ribosome as a mature protein.

5. The Tryptophan Operon

An operon is a DNA segment that includes an operator, a common promoter, and one or more structural genes. All of these structural genes are controlled as a single unit to produce messenger RNA (mRNA). Operons occur primarily in very simple organisms as prokaryotes and nematodes. The operon concept was introduced by François Jacob and Jacques Monod in 1961, though their studies on what is now known as the lactose operon.

A promoter is a short DNA sequence that provides a site for RNA polymerase to bind and initiate transcription; thus, it is located before the structural genes. Close to the promoter, and usually beside it, lies an operator sequence. An operator is a segment of DNA that regulates the activity of the operon promoter by interacting with a specific protein. This protein can act either as a repressor or as an activator. The operon may also contain regulatory genes, such as a repressor gene, which codes for a protein that binds to the operator and inhibits transcription.

Tryptophan (Trp) is one of the 20 main amino-acids in the genetic code (codon UGG). It is an essential amino acid because it cannot be synthesized by mammals, and therefore must be part of our day-to-day diet. Among other important substances, tryptophan is a precursor for serotonin (a neurotransmitter) and melatonin (a neurohormone).

Tryptophan can be synthesized by bacteria like *E. coli* through a series of catalysed reactions. The catalyzing enzymes in *E. coli* are made up of the polypeptides encoded by the tryptophan operon genes: *(trpE, trpD, trpC, trpB,* and *trpA*). These genes are transcribed from *trpE* to *trpA*. Finally, transcription is initiated at promoter *trpP*, which is indeed located just before the gene *trpE*.

The *trp* operon is regulated by three different negative-feedback mechanisms: repression, transcription attenuation, and enzyme inhibition. Below, these regulatory mechanisms are briefly reviewed. It is convenient for this to refer to Figure 2.

The *trp* operon is a repressible operon. This happens because there is an operator *trpO* overlapping with the operon promoter, *trpP*. When an active repressor is bound to *trpO* it blocks the binding of a mRNA to *trpP* and prevents transcription initiation. The *trp* repressor normally exists as a dimeric protein (called the *trp* aporepressor) and may or may not be complexed with tryptophan (Trp). Each portion of the *trp* aporepressor has a binding site for tryptophan.

When not complexed with tryptophan, the *trp* aporepressor cannot bind tightly to the operator *trpO*. However, if two tryptophan molecules bind to their respective binding sites, the *trp* aporepressor is converted into the functional repressor. The resulting functional repressor complex can bind tightly to the *trp* operator, and so the synthesis of tryptophan catalyzing enzymes is prevented. This fact completes the repression negative-feedback mechanism: An increase in the concentration of tryptophan induces an increase in the concentration of the functional repressor complexes, thus preventing the synthesis of tryptophan.

Transcription attenuation works by promoting an early termination of mRNA transcription, see Figure 3. The transcription starting site in the *trp* operon is separated from *trpE* by a leader region responsible for attenuation control. The transcript of this leader region consists of four segments which can form three stable hairpin structures between consecutive segments. The first segment contains two tryptophan codons in tandem. If there is an

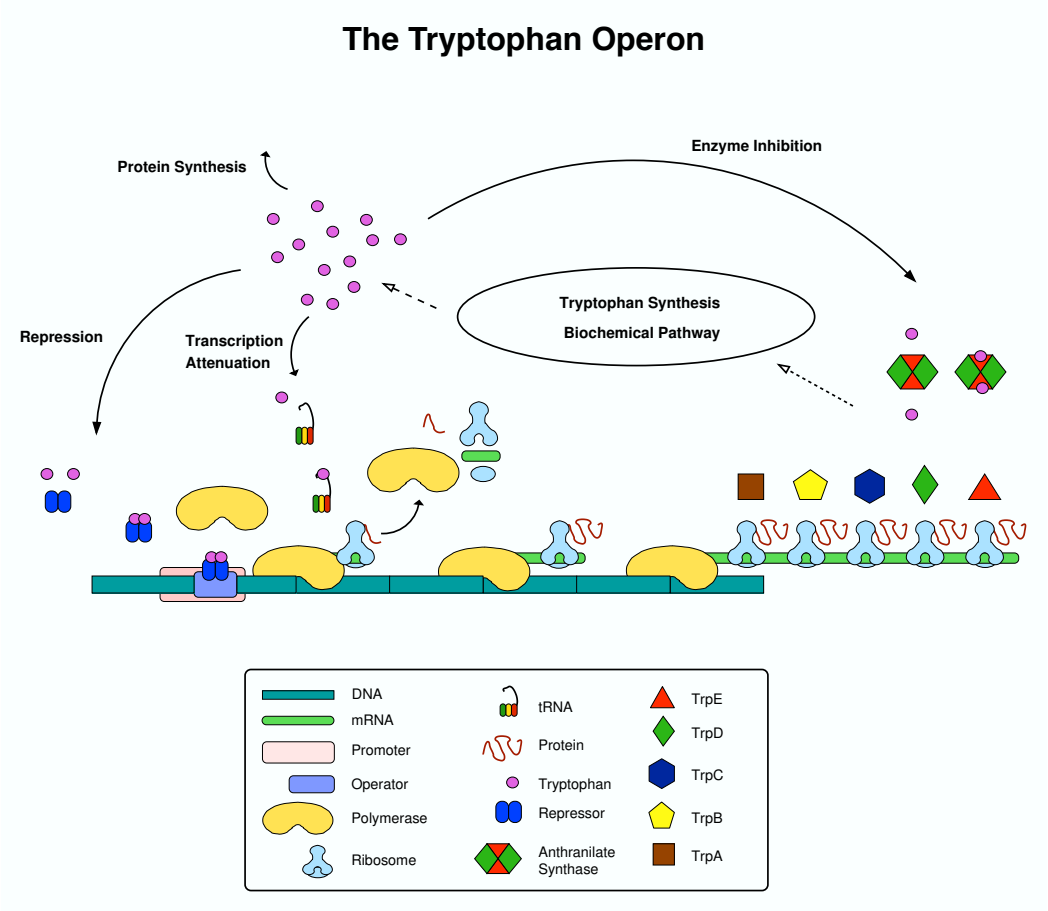

Figure 2. Schematic representation of the *trp* operon regulatory mechanisms.

abundance of tryptophan, and thus of loaded $tRNA^{Trp}$, the ribosome rapidly finishes translation of the first two segments, and so it promotes the formation of a stable hairpin structure between the last two segments. mRNA polymerase molecules recognize this hairpin structure as a termination signal, and transcription is prematurely terminated. However, if the ribosome stalls in the first segment due to lack of tryptophan, hairpin development between Segments 2 and 3 (the antiterminator) is facilitated, and transcription proceeds until the end.

Finally, anthranilate synthase is the first enzyme to catalyze a reaction in the catalytic pathway that leads to the synthesis of tryptophan from chorismate. This enzyme is a heterotetramer consisting of two TrpE and two TrpD polypeptides. Anthranilate synthase is inhibited by tryptophan by negative-feedback. This feedback inhibition is achieved when the TrpE subunits in anthranilate synthase are individually bound by a tryptophan molecule. Therefore, an excess of intracellular tryptophan inactivates most of the anthranilate synthase protein, avoiding the production of more tryptophan.

We recommend References [13, 14, 15] for those interested in reading more about the regulatory mechanisms in the *trp* operon.

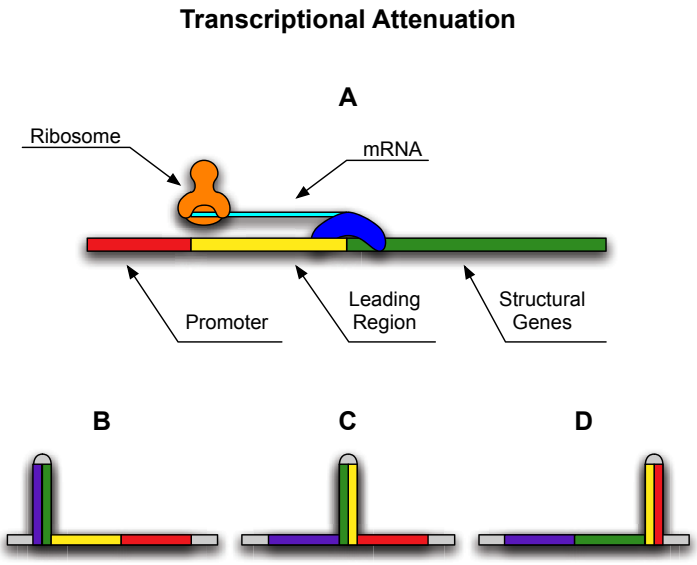

Figure 3. A) In the tryptophan operon there is a leading region between the promoter and the structural genes. The transcript (cyan) of the *trp* leading region comprises four equally large segments, which can fold to form three different hairpin structures. B) After Segments 1 (purple) and two (green) have been transcribed, they form a hairpin (the Hairpin 1:2) and it causes the polymerase to stop. A ribosome may then bind to the nascent mRNA and starts translation; transcription is resumed when the ribosome disrupts Hairpin 1:2. C) Segment 1 contains two Trp codons in tandem. Hence, under conditions of low tryptophan, there is a reduced number of loaded $tRNA^{Trp}$, and so the ribosome gets stacked in Segment 1. Transcription continues anyway, so that Segment 2 forms a hairpin with Segment 3 (yellow) when the latter is transcribed. Hairpin 2:3 is recognized as an antiterminator by the polymerase, and so transcription proceeds until the end of the structural genes. D) Conversely, if there is an abundance of tryptophan, the ribosome rapidly starts translation of Segment 2, and precludes formation of Hairpin 2:3. Then, when Segments 3 and 4 are transcribed, they form a hairpin that is recognized as a terminator; it destabilizes the polymerase-DNA complex and prematurely aborts transcription.

6. Mathematical Modelling

The tryptophan operon has been the object of intensive studies for more than fifty years. The detailed knowledge we have today regarding the regulatory mechanisms in this operon is impressive. We mainly owe this knowledge to the research carried out at Charles Yanofsky's lab at Stanford University. However, there are still some open questions concerning the dynamic tryptophan operon behaviour. One of this questions is why the tryptophan operon involves three, apparently redundant, negative feedback regulatory mechanisms: repression, transcriptional attenuation, and enzyme inhibition. Together with other groups, we have addressed this question from a mathematical modelling perspective to try to gain more

insight. The rest of this section describes our results.

Table 1. Mathematical model of the tryptophan operon. These equations govern the dynamic evolution of the concentration of mRNA, M; enzyme, E; and intracellular tryptophan, T. The positive terms on the right-hand side of the equations stand for the production rates of the corresponding variables, while the negative terms stand for loss due to dilution (due to cell growth) and degradation. The term $\rho T/(K_\rho + T)$ represents tryptophan consumption during the synthesis of all proteins. The constant k_M is the rate of transcription initiation per promoter, k_E is the rate of translation initiation per mRNA, and k_T is the rate of tryptophan production per enzyme E. The terms γ_M, γ_E and γ_T are the dilution plus degradation rates. Finally, the functions $\mathcal{R}_R(T)$, $\mathcal{R}_A(T)$, and $\mathcal{R}_I(T)$ are all nonlinear decreasing functions of T, and respectively represent the three different regulatory mechanisms present in the tryptophan operon: repression, transcription attenuation, and enzyme inhibition.

$$
\begin{aligned}
\frac{dM}{dt} &= k_M D \mathcal{R}_R(T) \mathcal{R}_A(T) - \gamma_M M, \\
\frac{dE}{dt} &= k_E M - \gamma_E E, \\
\frac{dT}{dt} &= k_T E \mathcal{R}_I(T) - \rho \frac{T}{K_\rho + T} - \gamma_T T, \\
\mathcal{R}_R(T) &= \frac{\dfrac{P}{K_P}}{1 + \dfrac{P}{K_P} + \dfrac{R}{K_R} + \left(\dfrac{T}{T + K_T}\right)^2}, \\
\mathcal{R}_A(T) &= \frac{1 + 2\alpha \dfrac{T}{K_G + T}}{\left(1 + \alpha \dfrac{T}{K_G + T}\right)^2}, \\
\mathcal{R}_I(T) &= \left(\frac{K_I}{T + K_I}\right)^2 .
\end{aligned}
$$

We have developed a mathematical model of the tryptophan operon regulatory pathways, which takes into account all three known regulatory mechanisms: repression, transcription attenuation, and enzyme inhibition. The model equations are presented in Table 1. These equations govern the dynamic evolution of the concentration of mRNA, M; enzyme, E; and intracellular tryptophan, T. The positive right-hand-side terms at the equations stand for the production rates of the corresponding variables, while the negative terms stand for loss due to dilution (due to cell growth) and degradation. The term $\rho T/(T + K_\rho)$ represents tryptophan consumption during the synthesis of all types of proteins. The constant k_M is the rate of transcription initiation per promoter, k_E is the rate of translation initiation per mRNA, and k_T is the rate of tryptophan production per enzyme E. The terms γ_M, γ_E, and γ_T are the dilution plus degradation rates. Finally, the functions $\mathcal{R}_R(T)$, $\mathcal{R}_A(T)$,

and $\mathcal{R}_I(T)$ are all nonlinear decreasing functions of T, and respectively represent the three different regulatory mechanisms present in the tryptophan operon: repression, transcription attenuation, and enzyme inhibition. To derive these functions we took into consideration all of the biochemical reactions underlying the regulatory mechanisms described above, used chemical kinetics, and made quasi-steady state assumptions for all fast processes. The meaning of the parameters in the functions $\mathcal{R}_R(T)$, $\mathcal{R}_A(T)$, and $\mathcal{R}_I(T)$ is as follows: P represents the intracellular mRNA polymerase (mRNAP) concentration, K_P is the dissociation constant for the mRNAP-promoter complex formation reaction, K_R is the dissociation rate for the repressor-operator complex formation reaction, K_T is the dissociation constant for the reaction in which a tryptophan binds one of its corresponding binding sites in the aporepressor, α is a constant associated to the strength of transcription attenuation, K_G is the dissociation rate for the tryptophan-tRNATrp complex formation reaction, and K_I is the dissociation rate for the reaction in which a tryptophan molecule binds one of its binding sites in the anthranilate synthase enzyme.

It is important to mention that, although not introduced here, special attention was given to the estimation of all the model parameters from reported experimental data. The estimated parameter values are tabulated in Table 2. The reader interested in the derivation of the model equations, as well as in the estimation of the model parameters, may consult Reference [16].

Table 2. The model parameters as estimated in Reference [16]

μ	$\approx$	$2.3 \times 10^{-2}\,\mathrm{min}^{-1}$	P	$\approx$	$1,500\,\mathrm{mpb}$	O	$\approx$	$2\,\mathrm{mpb}$
R	$\approx$	$400\,\mathrm{mpb}$	K_T	$\approx$	$20,000\,\mathrm{mpb}$	K_R	$\approx$	$0.1\,\mathrm{mpb}$
K_P	$\approx$	$22.5\,\mathrm{mpb}$	α	$\approx$	18.5	K_G	$\approx$	$2,500\,\mathrm{mpb}$
K_I	$\approx$	$2,050\,\mathrm{mpb}$	γ_M	$\approx$	$0.69\,\mathrm{min}^{-1}$	γ_E	$\approx$	$0.01\,\mathrm{min}^{-1}$
ρ	$\approx$	$1.2 \times 10^5\,\mathrm{mpb}$	K_ρ	$\approx$	$5,000\,\mathrm{mpb}$	τ_E	$\approx$	$1\,\mathrm{min}$
k_M	$\approx$	$5.1\,\mathrm{min}^{-1}$	k_E	$\approx$	$30\,\mathrm{min}^{-1}$	k_T	$\approx$	$3.2 \times 10^4\,\mathrm{min}^{-1}$

7. Dynamic Influence of the Three Regulatory Mechanisms in the *trp* Operon

Once we had the model, the next step was to analyze the dynamic influence of the three different regulatory mechanisms. To test the effect of enzyme inhibition, the model was modified to mimic a tryptophan operon in which enzyme inhibition is the only regulatory system, as well as a trp operon lacking enzyme inhibition. Then, we simulated derepression experiments, in which a bacterial culture that has grown for a long time in a medium rich in tryptophan (to shut the trp operon off) is suddenly shifted to a tryptophan-free medium (so the operon is reactivated).

After carrying out these simulations we observed that, when enzyme inhibition regulates Trp production by its own, the enzyme activity returns to its steady state almost

immediately [16]. Thus, this mechanism is quite efficient, from the viewpoint of control theory, in maintaining a steady state. Apparently, the reason for this high controlling efficiency is that when Trp concentration is high, tryptophan acts as a buffer which captures the enzymes that catalyze Trp production, and these enzymes are rapidly released when the Trp concentration decreases. On the other hand, in the simulations in which enzyme inhibition is absent, we observed that the operon produces enzymes at high rates, and this results in an overshoot of Trp production; approximately sixty times the production in normal conditions [17]. In our interpretation, this overshot is due to the much longer characteristic times of repression and attenuation. In conclusion, it seems that enzyme inhibition increases the operon stability because, due its rapid response, it relieves the system from the necessity to synthesize large amounts of polypeptide under conditions of tryptophan starvation.

It is known that regulation by transcription attenuation is exercised over the range from mild to extreme tryptophan depletion, while repression regulates over the range from excess tryptophan to mild Trp starvation. Thus, transcription attenuation increases the trp operon sensitivity to changes that alter the need for endogenous tryptophan. To investigate whether or not this system has any other dynamic effects, an operon reactivation simulation was carried out with a *trp* operon lacking transcription attenuation. The results were then compared with those of the normal operon. Our observations indicate that transcription attenuation makes the system reactivate sooner. Thus, this mechanism accelerates the trp operon response to nutritional shifts, by increasing its sensitivity range. In conclusion, enzyme inhibition and transcription attenuation provide the trp operon with important dynamic advantages. Enzyme inhibition increases the system stability, and transcription attenuation speeds up its response to nutritional shifts.

We further investigated the stability of the tryptophan operon model by means of the second Lyapunov method to generalize the results described previously [17]. First, we proved that the unique fixed point of the system is stable for a wide range of the parameters that determine the intensity of transcription attenuation and enzyme inhibition. Afterwards, we proceeded to analyze the stability strength in the *wild-type*, *inhibition-less* and *attenuation-less* bacterial strains. From this, we concluded, in agreement with the numerical results, that both regulatory mechanisms strengthen the system stability. Nevertheless, while the lack of enzyme inhibition greatly weakens the stability of the system fixed-point, the dynamic influence of transcription attenuation is much less important, since it speeds up the operon response but only slightly. In conclusion, enzyme inhibition is very important from a dynamic viewpoint. Conversely, the main effect of transcription attenuation is increasing the trp-operon sensitivity range to nutritional shifts, whereas its effect on the system stability is much weaker.

8. Comparison with Experimental Results and Model Improvement

As we said in the previous section, our *trp* operon model allowed us to study the dynamic influence of its three different regulatory mechanisms, and analyze them from an evolutionary perspective. However, we still need to test the model feasibility, by comparing it with reported experimental dynamic results, before we can be completely confident about

its predictions. In this respect, there are some experiments carried out by Yanofsky's group in which a bacterial culture that had grown for a long time in a medium rich in tryptophan —to shut the trp operon off— was suddenly shifted to a tryptophan-free medium, and the temporal evolution of the corresponding genes' expression level was measured [18]. These experiments will be referred to as derepression experiments. On the other hand, Bliss et al. [19] carried out derepression experiments with an *E. coli* mutant strain in which enzyme inhibition is attenuated. They observed that the phase-space trajectories do not converge to the steady state but oscillate in a limit cycle.

We simulated the derepression experiments of Yanofsky and Horn [18] and compare them with the model results in Figure 4A). Notice that, according to our model, the operon activity level should recover more slowly than it actually does. Moreover, we were also unable to reproduce the oscillatory behaviour observed by Bliss et al. by modifying the parameters corresponding to the enzyme inhibition regulatory function (to mimic their mutant *E. coli* straint). At this point, we could have started to arbitrarily modify the model parameters to fit the experimental results, but instead we decided to trust our estimations and wondered whether some important aspect in the biology had been neglected.

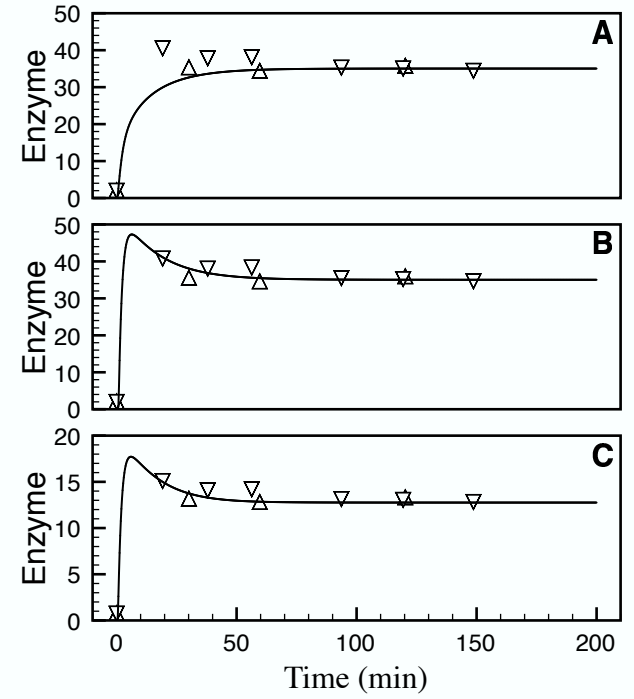

Figure 4. Plots of enzyme count *vs.* time resulting from derepression experiments simulated with the model introduced in [16] (A), with the same model but considering the time delays due to transcription and translation (B), and the modified model that considers the existence of three different operators, including the time delays (C).

Transcription and translation are not instantaneous processes. Compared with all the other biochemical processes involved in the *trp* operon regulatory pathways, they take long

times to occur. In a previous work [17] we estimated the time delays associated with transcription ($\tau_M \simeq 1\,\text{s}$) and translation ($\tau_E < 1\,\text{min}$), but dismissed them apriori because of their shortness. Here, we take them into consideration to explicitly test their effect on the system dynamics. To do that, we have to modify the differential equations governing the dynamics of M and E as follows:

$$\begin{aligned} \frac{dM}{dt} &= k_M D \mathcal{R}_R(T_{\tau_M}) \mathcal{R}_A(T_{\tau_M}) - \gamma_M M, \\ \frac{dE}{dt} &= k_E M_{\tau_E} - \gamma_E E, \end{aligned}$$

where the notation X_τ means that variable X is delayed a time τ, i.e. $X_\tau(t) = X(t-\tau)$. We then numerically solved these modified equations with the aid of the program `xppaut`. The results are plotted in Figure 4B, where we used $\tau_E = 12\,\text{s}$. Notice how, by simply taking into account such rather short time delays, there is a much better agreement between the model predictions and the experimental results of Yanofsky and Horn. On the other hand, despite this success, we were still unable to reproduce the oscillatory behaviour observed by Bliss et al. [19] by modifying the parameters associated to the enzyme inhibition regulatory function.

There is one further level of complexity in the *trp* operon regulatory pathway that we have not taken into account in our previous models. Namely the DNA regulatory region upstream of gene *trpE* contains three different repressor binding sites, denoted as O1, O2, and O3, and two repressors can cooperatively bind O1 and O2 [20]. This can be taken into account by modifying the function $\mathcal{R}_R(T)$ as follows (see the Appendix):

$$\mathcal{R}_R(T) = \frac{\frac{P}{K_P}}{\left(1+\frac{R_{2T}}{K_R^1}\right)\left(1+\frac{R_{2T}}{K_R^2}\right)\left(1+\frac{R_{2T}}{K_R^3}\right) + \frac{R_{2T}^2}{K_R^1 K_R^2}\left(1+\frac{R_{2T}}{K_R^3}\right)\left(k_{\text{cop}}-1\right) + \frac{P}{K_P}},$$

where

$$R_{2T} = R_{\text{Tot}}\left(\frac{T}{T+K_T}\right)^2$$

is the amount of active repressor, K_R^i ($i = 1, 2, 3$) is the dissociation constant for the R_{2T}-Oi complex formation reaction, and $k_{\text{cop}} > 1$ is a constant accounting for the cooperativity between Operators O1 and O2. The parameters for this new repression function are also estimated in the Appendix, and their values are as follows:

$$\begin{aligned} &K_R^1 = 0.625\,\text{mpb}, \quad K_R^2 = 7.9\,\text{mpb}, \quad K_R^3 = 100.0\,\text{mpb}, \\ &K_P = 2\,\text{mpb}, \qquad k_{\text{cop}} = 11.1125. \end{aligned}$$

All the other parameters remain set at the values estimated in Table 2.

After modifying the model to take into account the existence of three operators and the cooperativity observed between Operators O1 and O2, we simulated the derepression experiments of Yanofsky and Horn [18] with a time delay $\tau_E = 6\,\text{s}$; the results are shown in Figure 4C. Notice that, again, there is an excellent agreement between the model results and the experimental data.

As mentioned above, Bliss et al. [19] carried out derepression experiments with an *E. coli* mutant strain in which the enzyme anthranilate synthase cannot be inhibited by

tryptophan, and observed an oscillatory expression of the *trp* operon genes. After modifying the model to account for the time delays due to transcription and translation and the three repressor binding sites, we were able to reproduce the oscillations observed by Bliss et al.. For this, we reset the following parameters

$$K_I = 5 \times 10^7 \,\text{mpb} \quad \text{and} \quad K_\rho = 5,000 \,\text{mpb}.$$

The results of these new simulations are plotted in Figure 5. Although there is a good qualitative agreement between the model simulations and the model results of Bliss et al., there are still some important discrepancies. Namely, we obtained a longer oscillation period and we had to not only increase the value parameter K_I (which corresponds to an atrophied enzyme inhibition), but also to reduce the value of K_ρ (which implies a reduced catalytic efficiency of anthranilate synthase) to make the system oscillate. As far as we know, this last modification is not supported by the available experimental evidence on the effect of the mutation on the enzyme activity. We conclude from this that, most probably, there are still some important aspects of repression which are not accounted for in the improved model.

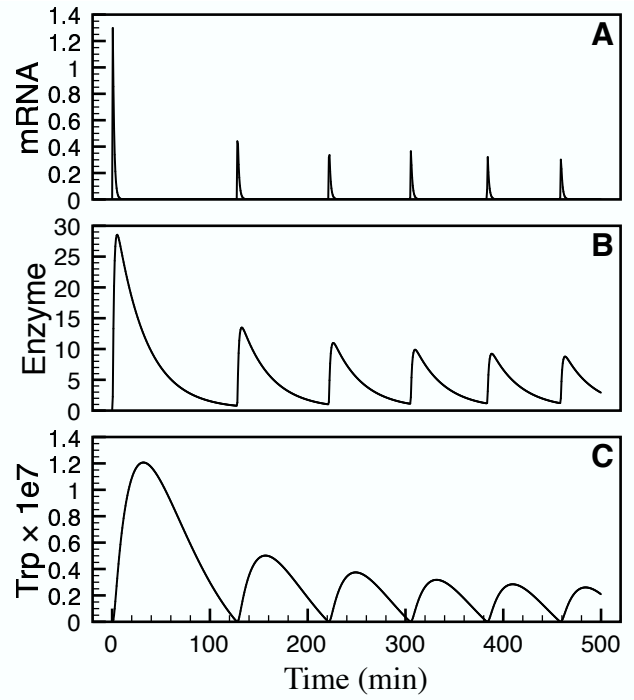

Figure 5. Plots of mRNA (A), enzyme (B), and tryptophan (C) molecule count resulting from derepression experiments simulated with the modified version of the model, which takes into account the time delays due to transcription and translation and the existence of three repressor binding sites.

9. Conclusions

We have reviewed out past efforts to understand the dynamic behaviour of the *trp* operon from a mathematical modelling approach. To develop the model we took into account the three different known mechanisms in the operon regulatory pathway. Special attention was paid to the estimation of the model parameters from reported experimental data. Among other things, the model was used to analyze the influence of the three different regulatory mechanisms (in the *trp* operon) over the system dynamic behaviour. These studies suggest that the system has three redundant negative-feedback regulatory mechanisms to guarantee a rapid response to variation on the growing medium. On the other hand, as seen in Figure 4A, the agreement of the model simulations with the experiments of Yanofsky and Horn [18] is rather poor. Besides, it is impossible to reproduce the oscillatory behaviour observed by Bliss et al. while carrying out derepression experiments with mutant *E. coli* strain in which anthranilate synthase cannot be inhibited.

Given that our models take into account the chemical details of all the regulatory mechanisms, and that all the parameters were estimated from experimental results, we interprete the consistent disagreement observed between the model results and the experiments of Yanofsky and Horn as a deficiency in the model. Then, we looked in the literature and found that there are indeed some important aspects which we did not consider. Namely there are three repressor binding sites, two of which interact cooperatively, and the processes of transcription and translation involve non-negligible time delays. Once the model was modified accordingly, it could reproduce the experiment results of Yanofsky and Horn, as well as the oscillatory data observed by Bliss et al..

An interesting conclusion, arising from the results discussed in the previous paragraph, is that the time delays have a strong influence on the system transient behaviour, despite their rather small value. In our opinion this is important because small time delays are usually neglected by employing the argument that they do not have an important effect on the dynamic system behaviour. If we were only interested on the system stationary behaviour, we could indeed neglect short time delays because only the long ones can cause bifurcations. However, if the transients are biologically meaningful (as in this case), no time delay should be ignored, regardless its value.

Finally, these results and conclusions reveal, in our opinion, the importance of developing detailed models whenever the biological information is available. With them, it is possible to gain a deeper insight into the system dynamics than with simpler phenomenological models.

A. Modelling the Three *trp* Operators

Repression Function

Up to three active repressors, R_{2T}, and one polymerase, P, can bind to the *trp* DNA regulatory region through the following reactions:

$$R_{2T} + O \overset{K_R^1}{\leftrightharpoons} O_R^1, \qquad R_{2T} + O \overset{K_R^2}{\leftrightharpoons} O_R^2, \qquad R_{2T} + O \overset{K_R^3}{\leftrightharpoons} O_R^3,$$

$$R_{2T} + O_R^1 \overset{K_R^2/k_{\text{cop}}}{\leftrightharpoons} O_{2R}^{12}, \quad R_{2T} + O_R^1 \overset{K_R^3}{\leftrightharpoons} O_{2R}^{13}, \quad R_{2T} + O_R^2 \overset{K_R^3}{\leftrightharpoons} O_{2R}^{23},$$

$$R_{2T} + O_{2R}^{12} \overset{K_R^3}{\leftrightharpoons} O_{3R}^{123}. \qquad P + O \overset{K_P}{\leftrightharpoons} O_P,$$

These these are not all the possible chemical reactions. However, the equilibrium equations arising from them, plus the conservation equation for the *trp* DNA regulatory region, form a complete system and therefore there is no need to consider more reactions. In the above reactions, O, O_R^i, O_{2R}^{ij}, O_{3R}^{123}, and O_P respectively denote the states in which all operators are free, in which only Operator Oi is bound by an active repressor, in which Operators Oi and Oj are both bound by active operators, in which all three operators are repressor-bounded, and in which a polymerase is bound to the promoter. Furthermore, K_R^i and K_P are the dissociation constants for the Oi:R_{2T} and promoter-polymerase complex formation reactions, respectively. Finally, $k_{\text{cop}} > 1$ denotes the cooperativity between Operators O1 and O2.

The equilibrium equations for these reactions are:

$$R_{2T}O = K_R^1 O_R^1, \qquad R_{2T}O = K_R^2 O_R^2, \qquad R_{2T}O = K_R^3 O_R^3,$$
$$R_{2T}O_R^1 = K_R^2 O_{2R}^{12}/k_{\text{cop}}, \quad R_{2T}O_R^1 = K_R^3 O_{2R}^{13}, \quad R_{2T}O_R^2 = K_R^3 O_{2R}^{23},$$
$$R_{2T}O_{2R}^{12} = K_R^3 O_{3R}^{123}.$$

From these equations, and the conservation equation for the total concentration of the *trp* DNA regulatory region, we have:

$$O + O_R^1 + O_R^2 + O_R^3 + O_{2R}^{12} + O_{2R}^{13} + O_{2R}^{23} + O_{3R}^{123} + O_P = O_{\text{Tot}},$$

It then follows that:

$$\begin{aligned} \mathcal{R}_R(T) &= \frac{O_P}{O_{\text{Tot}}} \\ &= \frac{\frac{P}{K_P}}{\left(1 + \frac{R_{2T}}{K_R^1}\right)\left(1 + \frac{R_{2T}}{K_R^2}\right)\left(1 + \frac{R_{2T}}{K_R^3}\right) + \frac{R_{2T}^2}{K_R^1 K_R^2}\left(1 + \frac{R_{2T}}{K_R^3}\right)(k_{\text{cop}} - 1) + \frac{P}{K_P}}. \end{aligned}$$

Parameter Estimation

According to [20], an active repressor molecule R_{2T} can bind to three different operator sites (O1, O2, and O3) all of which overlap the *trp* promoter. Additionally, two repressors bound to O1 and O2 interact in such a way that the corresponding binding energy is larger than the sum of the binding energies of single repressors separately binding to O1 and O2. Grillo et al. also report several measurements from which the following dissociation constants can be estimated:

$$\begin{aligned} K_R^1 &\simeq 0.625\,\text{mpb}, \\ K_R^2 &\simeq 7.9\,\text{mpb}, \\ K_R^3 &\simeq 100\,\text{mpb}. \end{aligned}$$

K_R^i denotes the dissociation constant for the reaction in which a repressor molecule binds the operator Oi alone. Since Grillo et al. also measured the binding energy when Operators O1 and O2 are simultaneously bound by repressor molecules, the following cooperativity constant can also be estimated from their experiments:

$$k_{\text{cop}} \simeq 11.13.$$

Finally, we used the fact that the operon activity decreases 60 times due to repression when there is abundance of tryptophan in the growing medium [13] (that is, $\mathcal{R}(0)/\mathcal{R}(T^*_{\text{max}}) \simeq 60$) to estimate parameter K_P:

$$K_P \simeq 5,000\,\text{mpb}.$$

T^*_{max} is the maximum steady-state intracellular concentration of tryptophan, and according to Bliss et al. [19] it is

$$T^*_{\text{max}} \simeq 20,000\,\text{mpb}.$$

Acknowledgments

This work was partially supported by *Consejo Nacional de Ciencia y Tecnología* (CONACyT), México, the Natural Sciences and Engineering Research Council (NSERC) of Canada, and the Canadian NCE Mathematics of Information Technology and Complex Systems (MITACS).

References

[1] J. D. Bernal. *Science in history*. MIT Press, Cambridge, Mass., 1971.

[2] I. Asimov. *Asimov's biographical encyclopedia of science and technology*. Doubleday, Garden City, NY, 2nd/rev edition, June 1982.

[3] D. Depew. *Darwinism evolving : systems dynamics and the genealogy of natural selection*. MIT Press, Cambridge, Mass., 1995.

[4] J. E. Cohen. Mathematics if biology's next microscope, only better; biology is mathematics next physics, only better. *PLoS Biology*, **2**(12):e439–e440, December 2004.

[5] H. V. Westerhoff and B. O. Palsson. The evolution of molecular biology into systems biology. *Nat. Biotechnol.*, **22**(10):1249–1252, October 2004.

[6] M. C. Mackey and M. Santillán. Mathematics, biology, and physics: interactions and interdependence. *Notices of the AMS*, **52**(8):832–840, September 2005.

[7] T. Ideker, T. Galitski, and L. Hood. A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.*, **2**:343–372, 2001.

[8] Y. Lazebnik. Can a biologist fix a radio?—Or,what I learned while studying apoptosis. *Cancer Cell*, **2**:179–182, 2002.

[9] A. Marshall. End of the interlude? *Nature Biotechnol.*, **22**:1191, 2004.

[10] S. Bornholdt. Less is more in modeling large genetic networks. *Science*, **310**:449–450, 2005.

[11] A. Friboulet and D. Thomas. Systems biology — an interdisciplinary approach. *Biosens. Bioelectron.*, **20**:2404–2407, 2005.

[12] M. W. Kirschner. The meaning of systems biology. *Cell*, **121**:503–504, May 2005.

[13] C. Yanofsky and I. P. Crawford. The tryptophan operon. In F. C. Neidhart, J. L. Ingraham, K. B. Low, B. Magasanik, and H. E. Umbarger, editors, *Escherichia coli and Salmonella thyphymurium: Cellular and Molecular Biology, Vol. 2*, pages 1453–1472. Am. Soc. Microbiol., Washington, DC, 1987.

[14] C. Yanofsky. Transcription attenuation, once viewed as a novel regulatory strategy. *J. Bacteriol.*, **182**:1–8, 2000.

[15] G. Xie, N. O. Keyhani, C. A. Bonner, and R. A. Jensen. Ancient origin of the tryptophan operon and the dynamics of evolutionary change. *Microbiol. Mol. Biol. Rev.*, **67**:303–342, 2003.

[16] M. Santillán and E. S. Zeron. Dynamic influence of feedback enzyme inhibition and transcription attenuation on the tryptophan operon response to nutritional shifts. *J. Theor. Biol.*, **231**:287–298, 2004.

[17] M. Santillán and E. S. Zeron. Analytical study of the multiplicity of regulatory mechanisms in the tryptophan operon. *Bull. Math. Biol.*, **68**:343–359, 2006.

[18] C. Yanofsky and V. Horn. Role of regulatory features of the trp operon of *Escherichia coli* in mediating a response to a nutritional shift. *J. Bacteriol.*, **176**:6245–6254, 1994.

[19] R. D. Bliss, P. R. Painter, and A. G. Marr. Role of feedback inhibition in stabilizing the classical operon. *J. Theor. Biol.*, **97**:177–193, 1982.

[20] A. O. Grillo, M. P. Brown, and C. A. Royer. Probing the physical basis for *trp* repressor-operator recognition. *J. Mol. Biol.*, **287**:539–554, 1999.